

# An introduction to univariate AR lag-1 state-space models for time-series data

Lecture 2  
EE Holmes

# Why are AR model, aka random walk?

- Lot's of complex processes can be approximated by AR models.

Today =  $f(\text{Yesterday}) + \text{noise}$

\* animal movement

\* gene frequency

Today = "noise"  $\times f(\text{Yesterday})$  take log  $\rightarrow$

$\log(\text{Today}) = \log(f(\text{Yesterday})) + \text{noise}$

\* population growth

# Topics

## Introduction to univariate AR lag-1 state-space models

- Definition of process versus observation error
- Examples of univariate random walks
- Hands on with some R code and simulations.
- Adding density dependence (feedbacks)

## Computer Labs

- Population viability analysis using state-space models
- Analysis of turtle track data (movement)

# Definitions: state-space

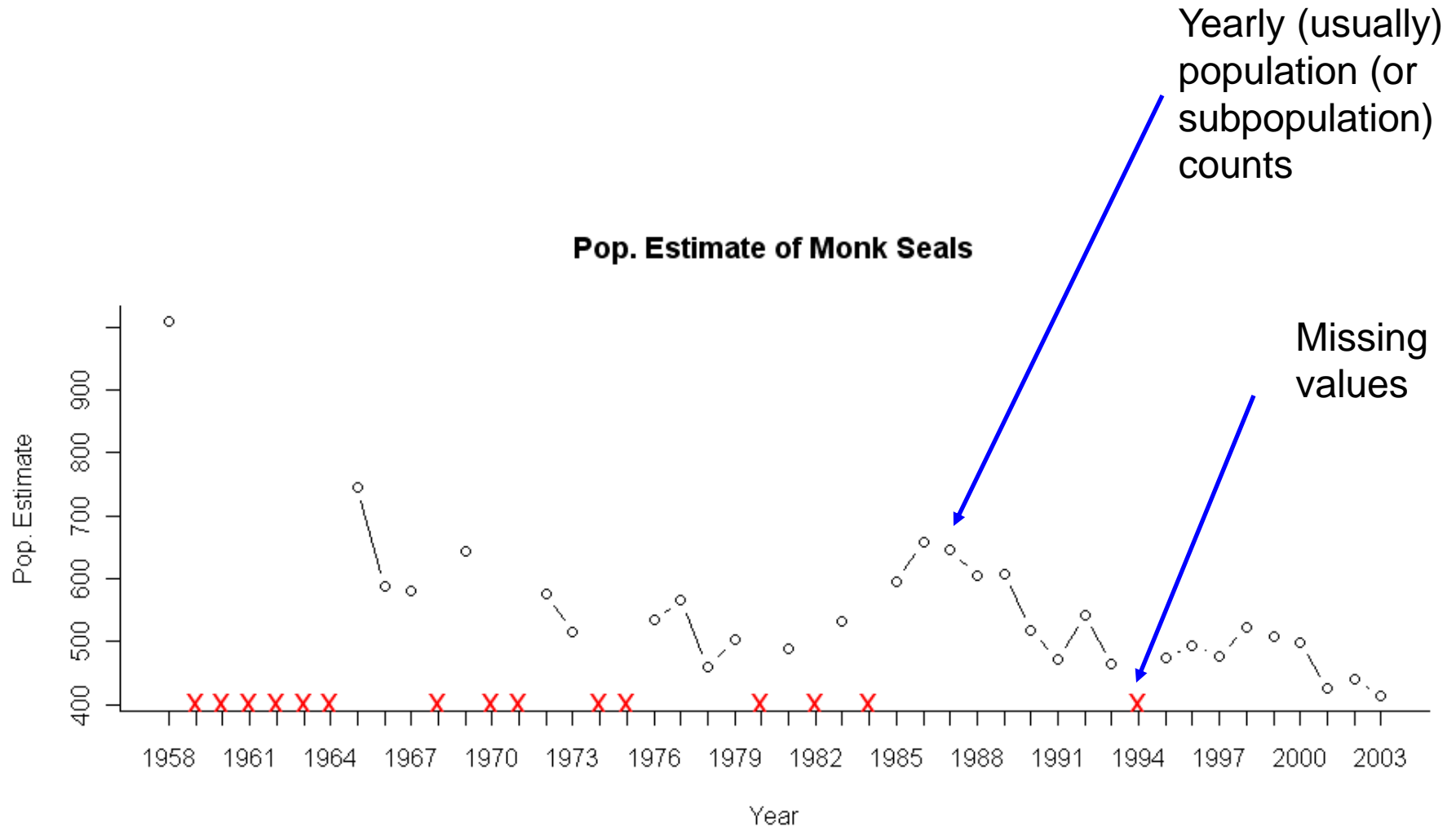
## Hierarchical statistical model

- "model within a model within a model ..."
  - $\text{data} = B \times \text{predictor}$  (a model for data)
  - $B = \text{normal}(\text{mean}=a, \text{s.d.}=\textcolor{red}{b})$  (a model for  $B$ )
  - $a = \text{binomial}(\textcolor{red}{p})$  (a model for parameter  $a$ )

## State-space model

- A special (and common) type of hierarchical model
  - $\text{data} = \text{function of } X(t) \times \text{some predictors (knowns)}$
  - $X(t)$  is a stochastic process with some parameters
  - In a MAR state-space model,  $X(t)$  is a *random walk*

# Example of univariate data: Count data

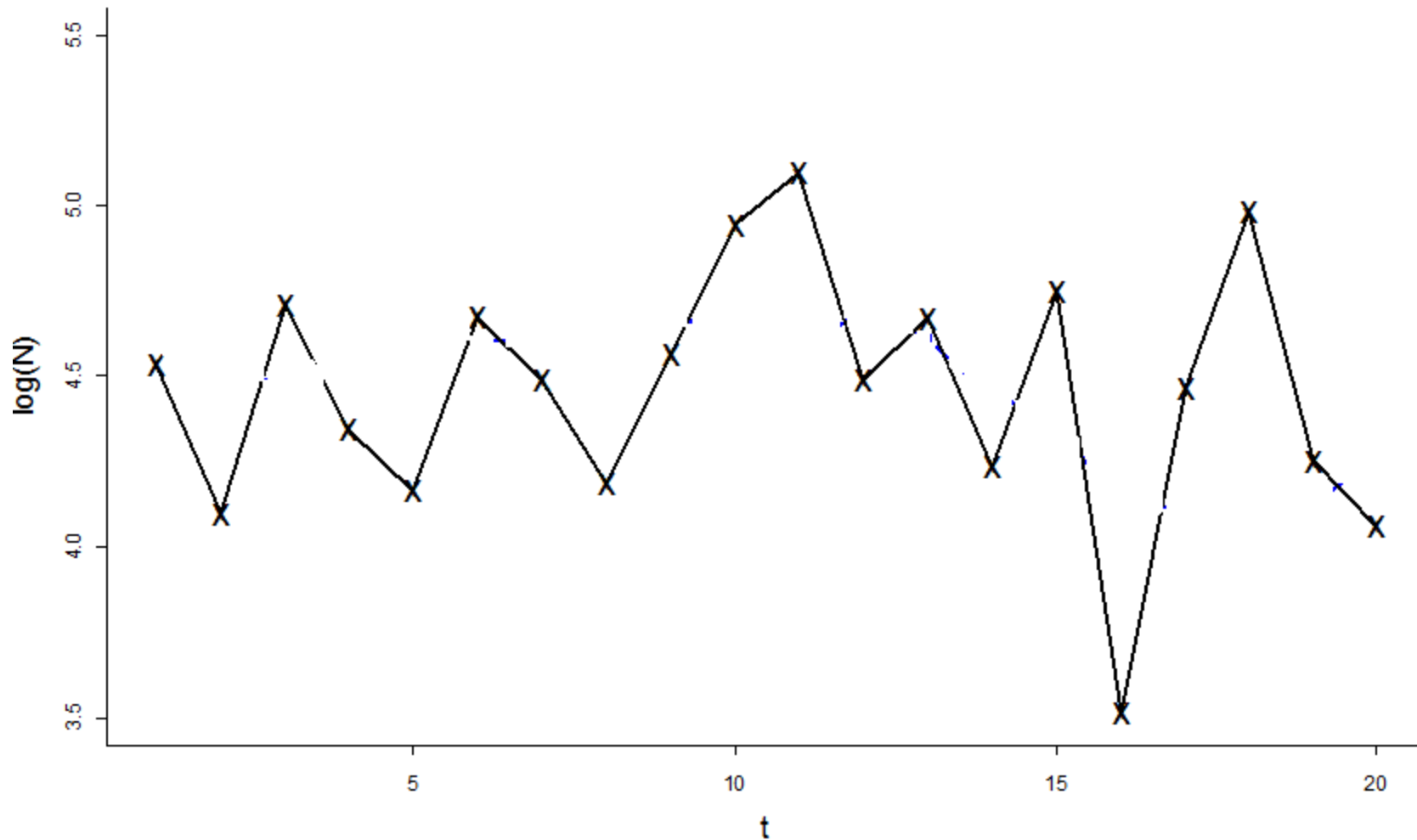


# Observation error

There **IS** some number of sea lions in our population in year  $x$ , but we don't know that number precisely. It is "hidden".

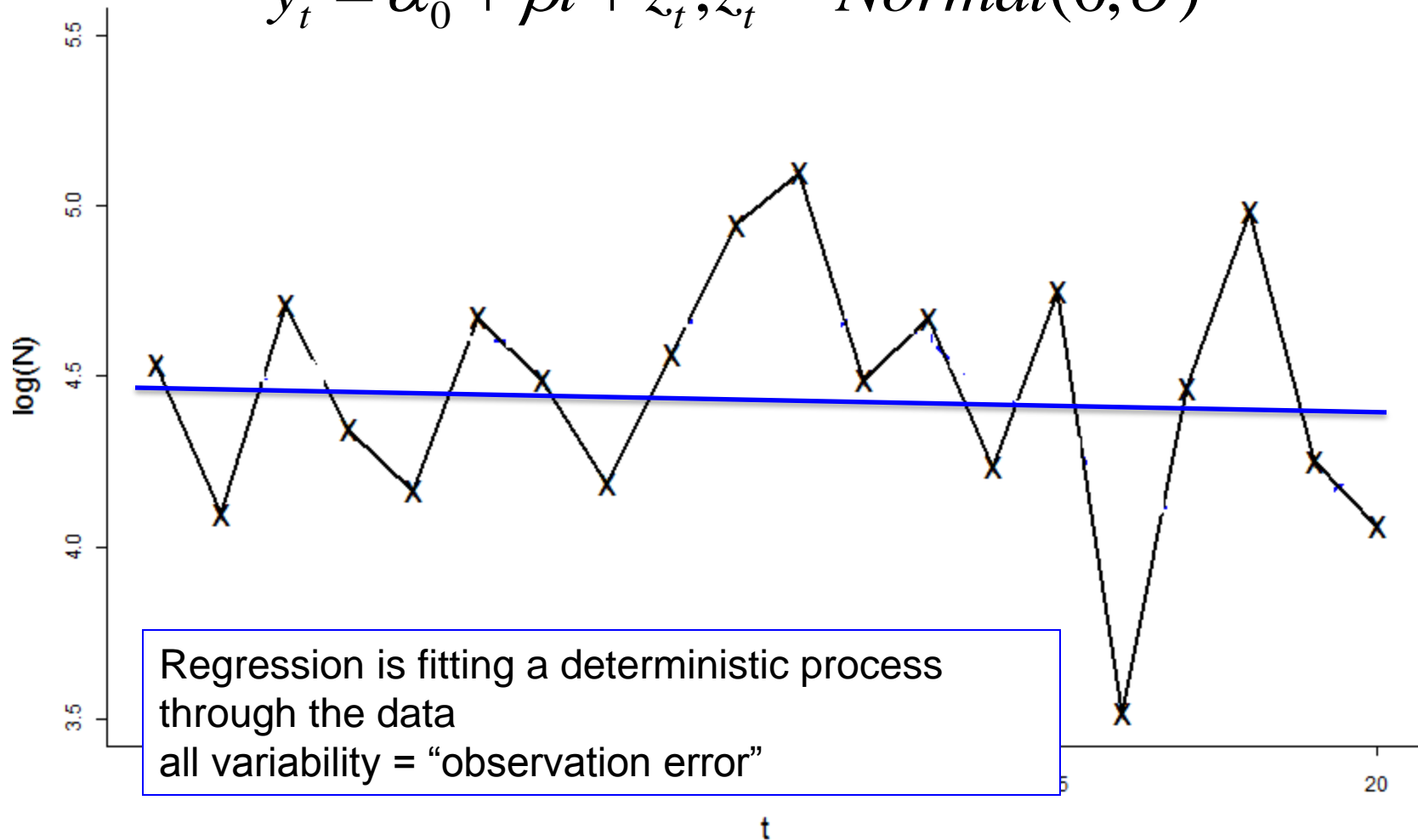


Suppose we have some count data (logged).

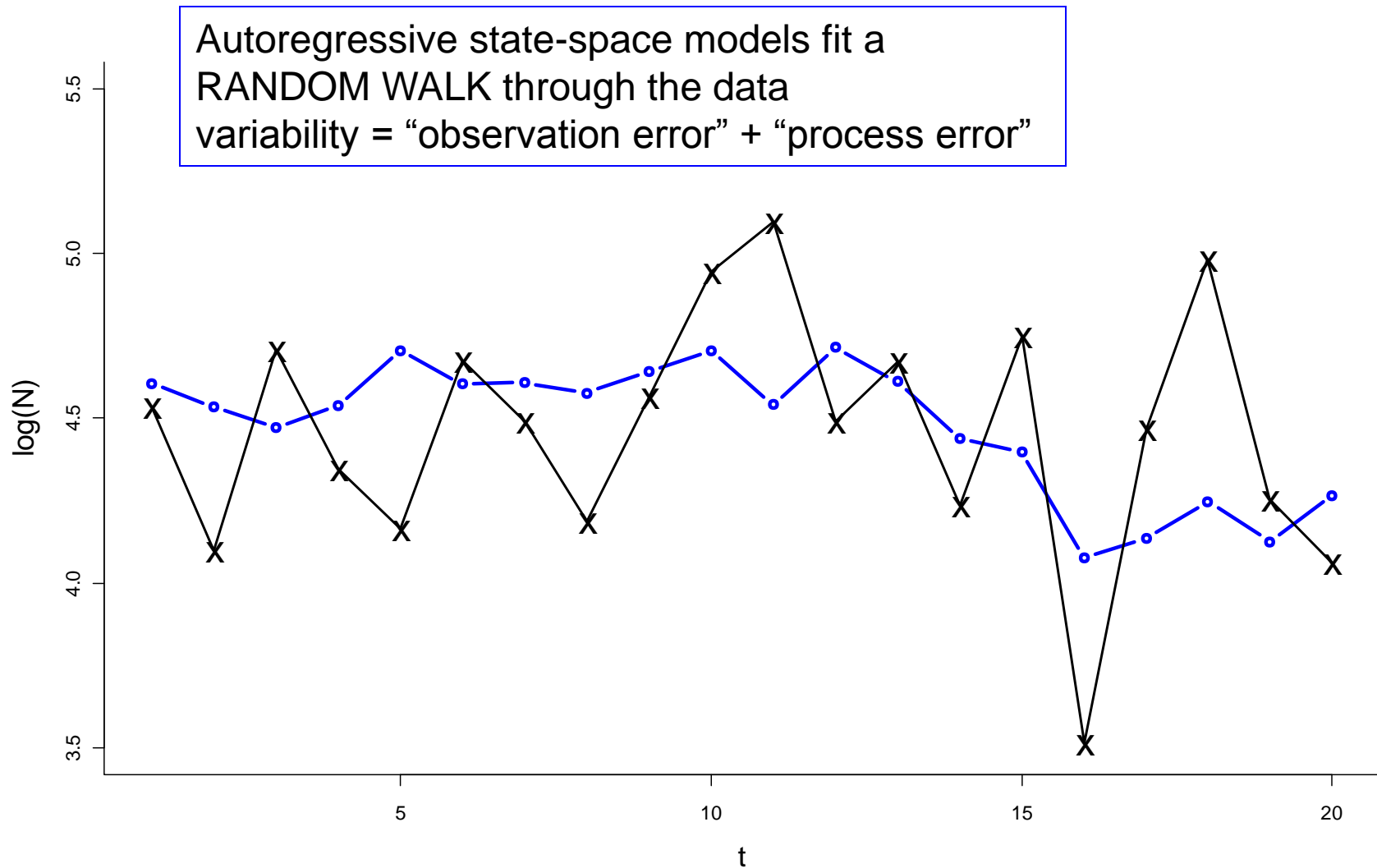


# What about fitting a regression line through the data?

$$y_t = \alpha_0 + \beta t + z_t; z_t \sim \text{Normal}(0, \sigma)$$

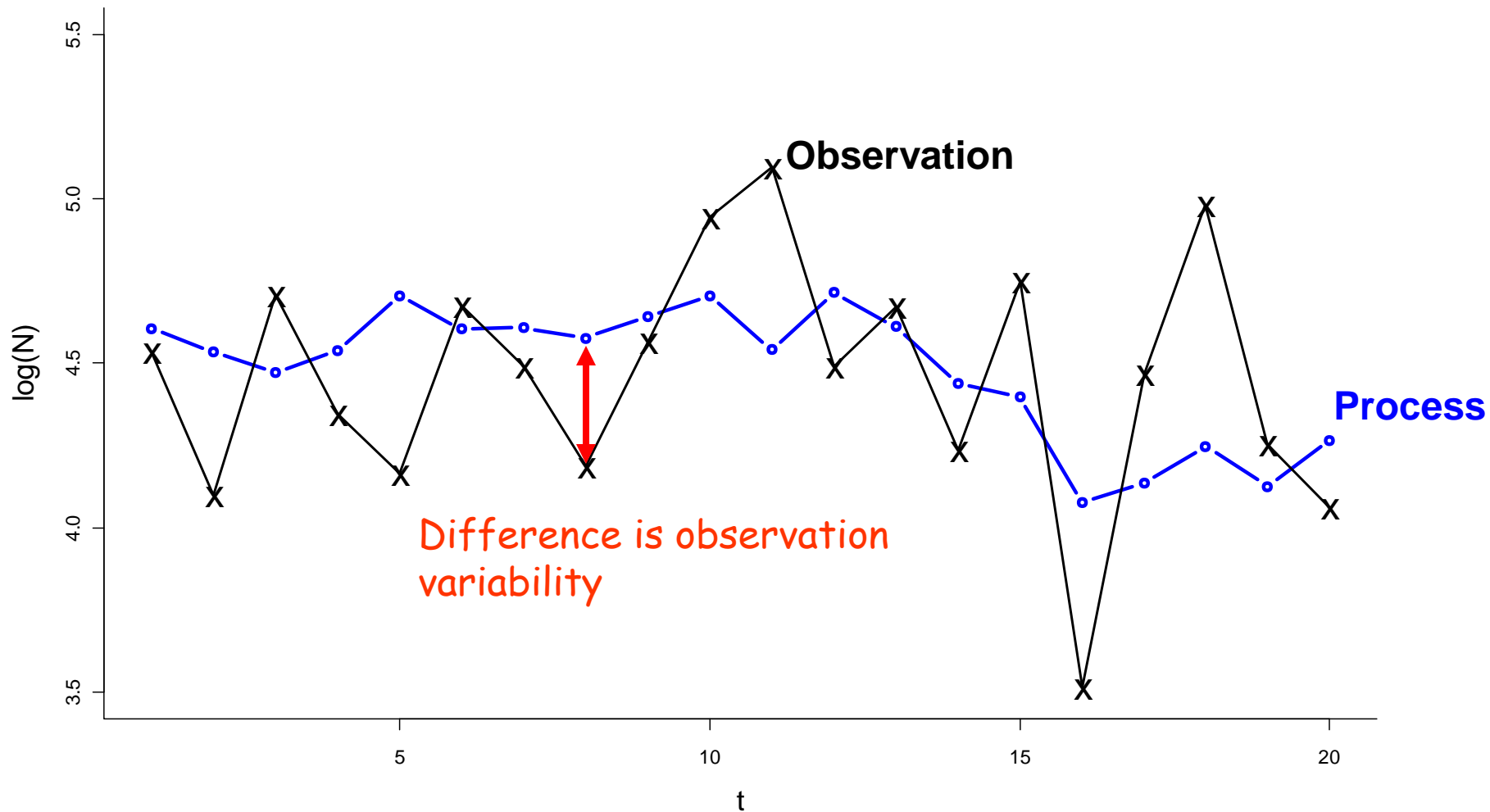


# Versus fitting an autoregressive state-space model



# Two types of variability

## #1 observation or “non-process” variability

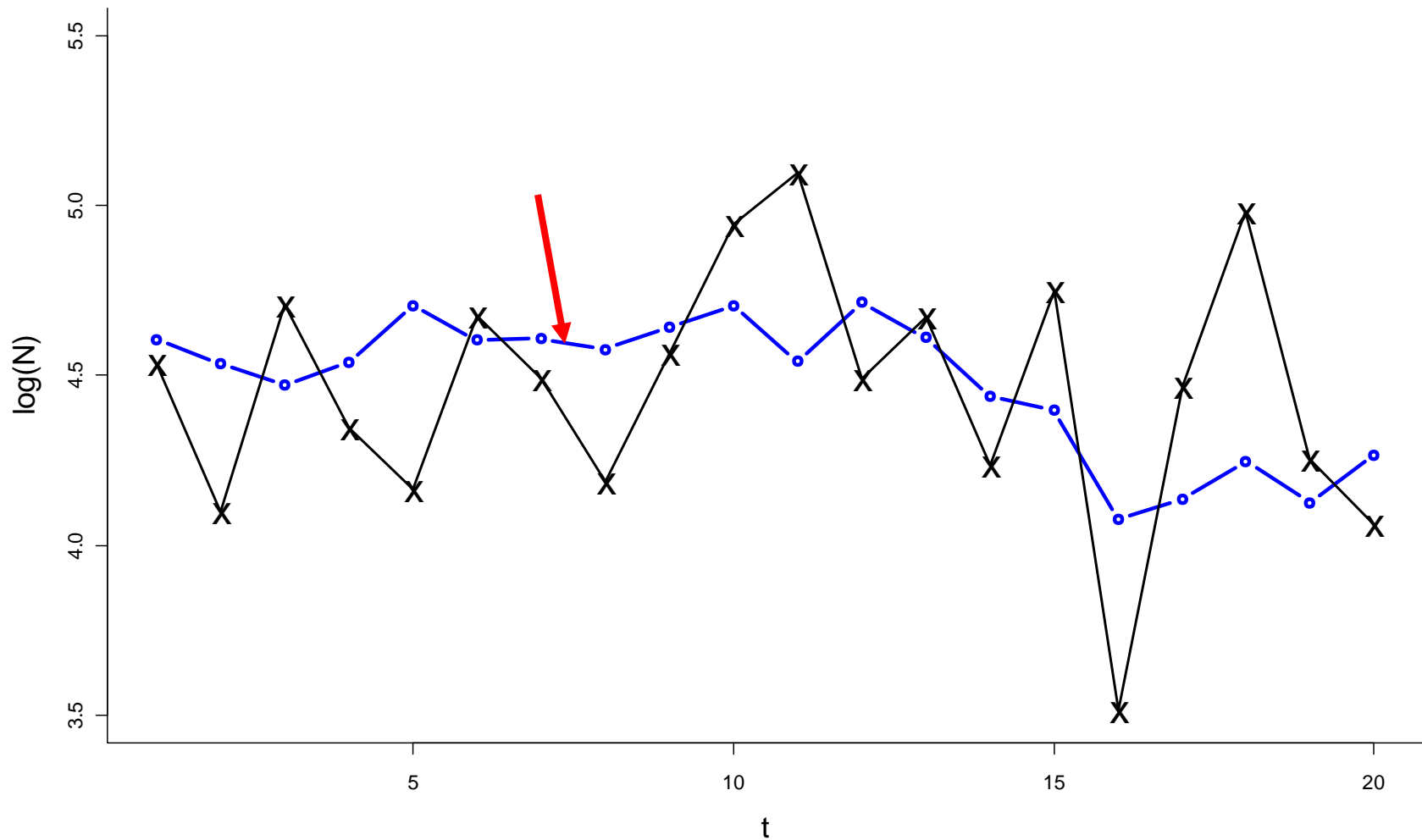


# The observation variance (and bias) is often unknowable

- Sightability varies (year-to-year, day-to-day, etc.) due to a myriad of factors that may not be fully understood or measureable
  - Environmental factors (tides, temperature, etc.)
  - Population factors (age structure, sex ratio, etc.)
  - Species interactions (prey distribution, prey density, predator distribution or density, etc.)
- Sampling variability--due to how you actually count animals--is just one component of observation variance

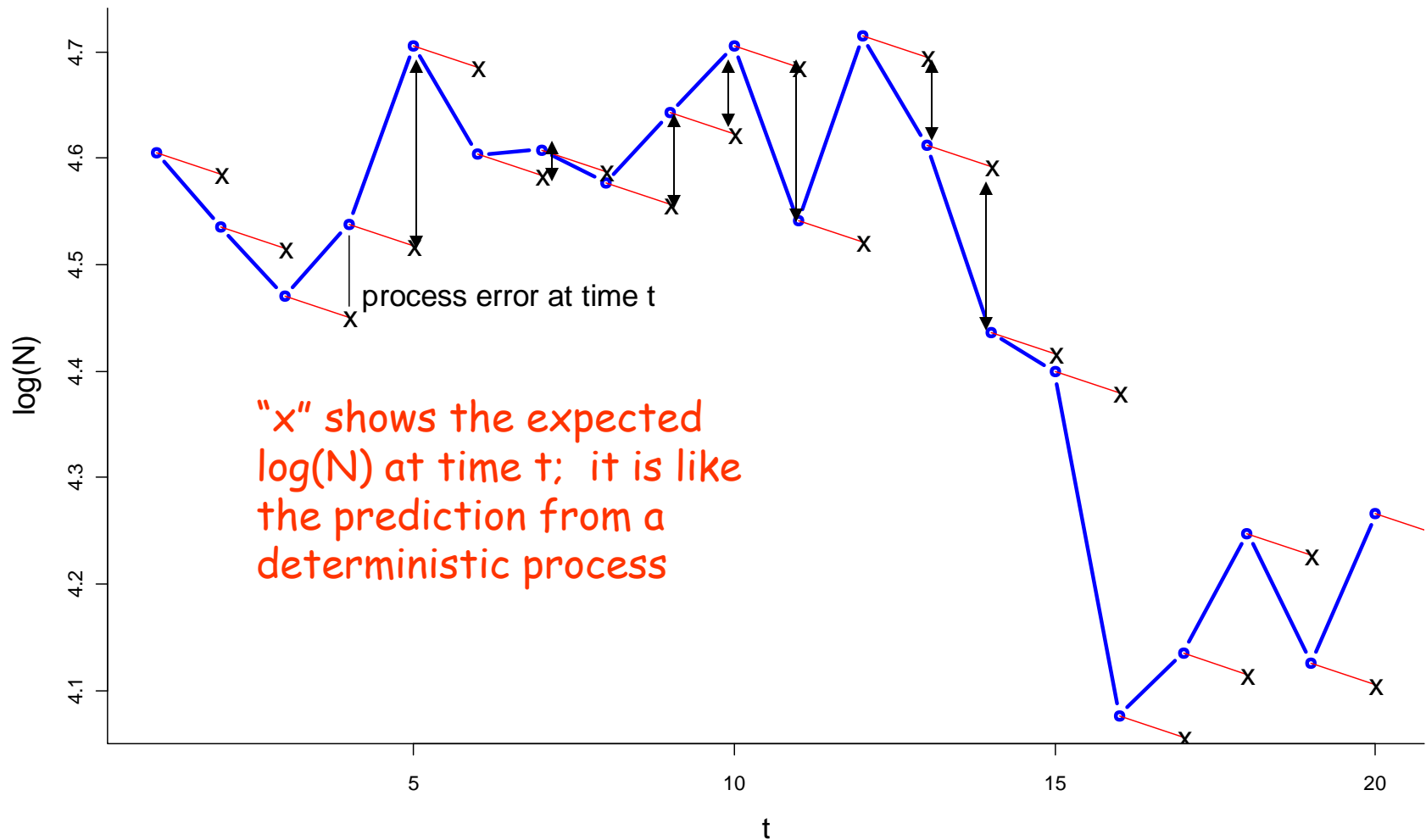
# Two types of variability

## #2 Process variability



# Process error is the difference between the expected population size and the actual value

Let's say that the mean rate of decline is 2% per year\*

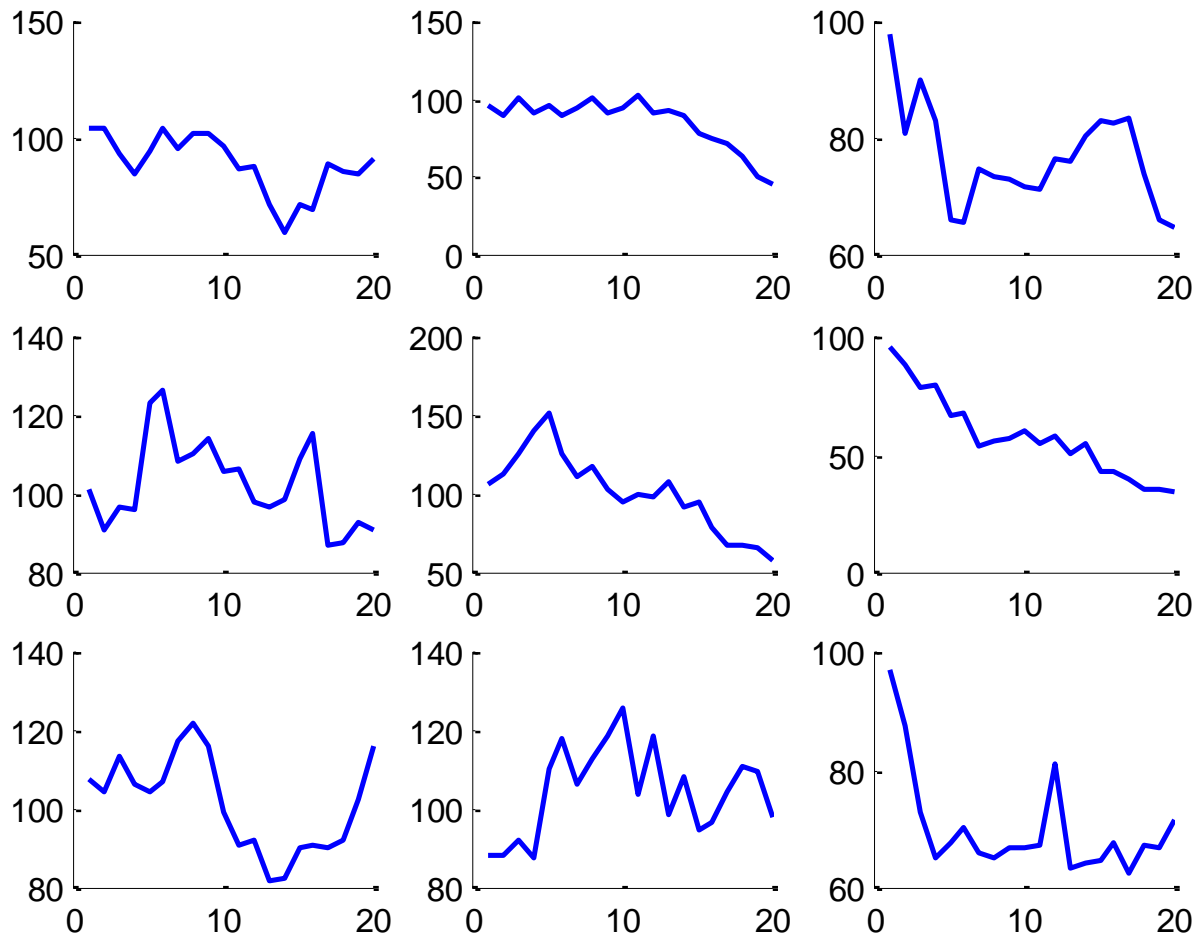


$$*u = -0.02$$

# The process error leads to characteristic random walks: AR lag-1 with drift

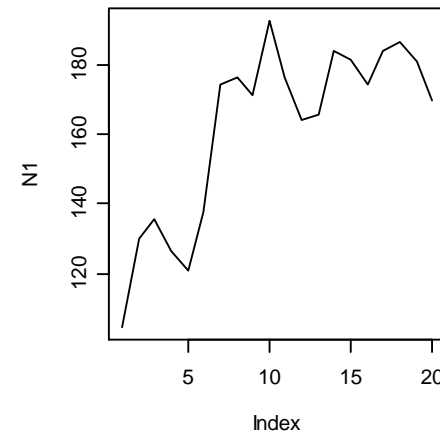
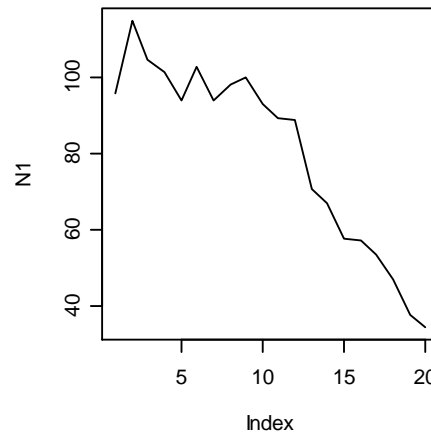
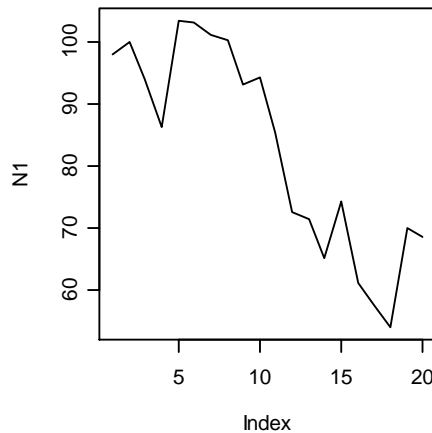
All trajectories came from the same model:

$$N_t = N_{t-1} \exp(-0.02 + e_t), \quad e_t \text{ was Normal}(\text{mean}=0.0, \text{var}=0.01)$$

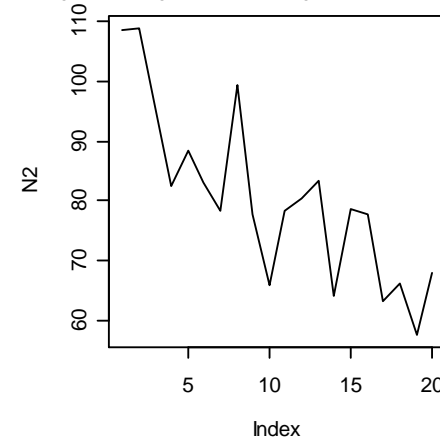
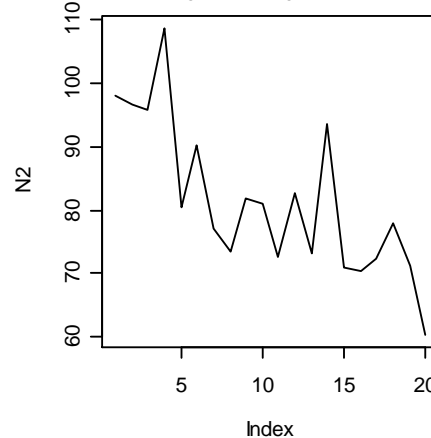
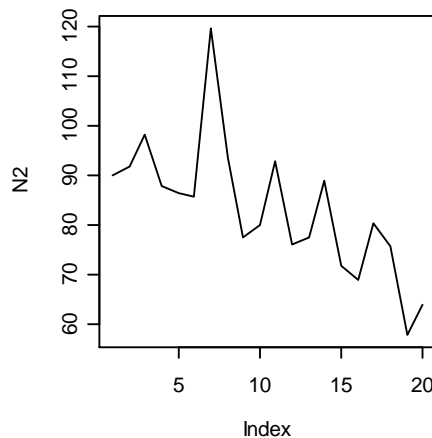


# How can we separate process and observation variance? They affect a time series differently.

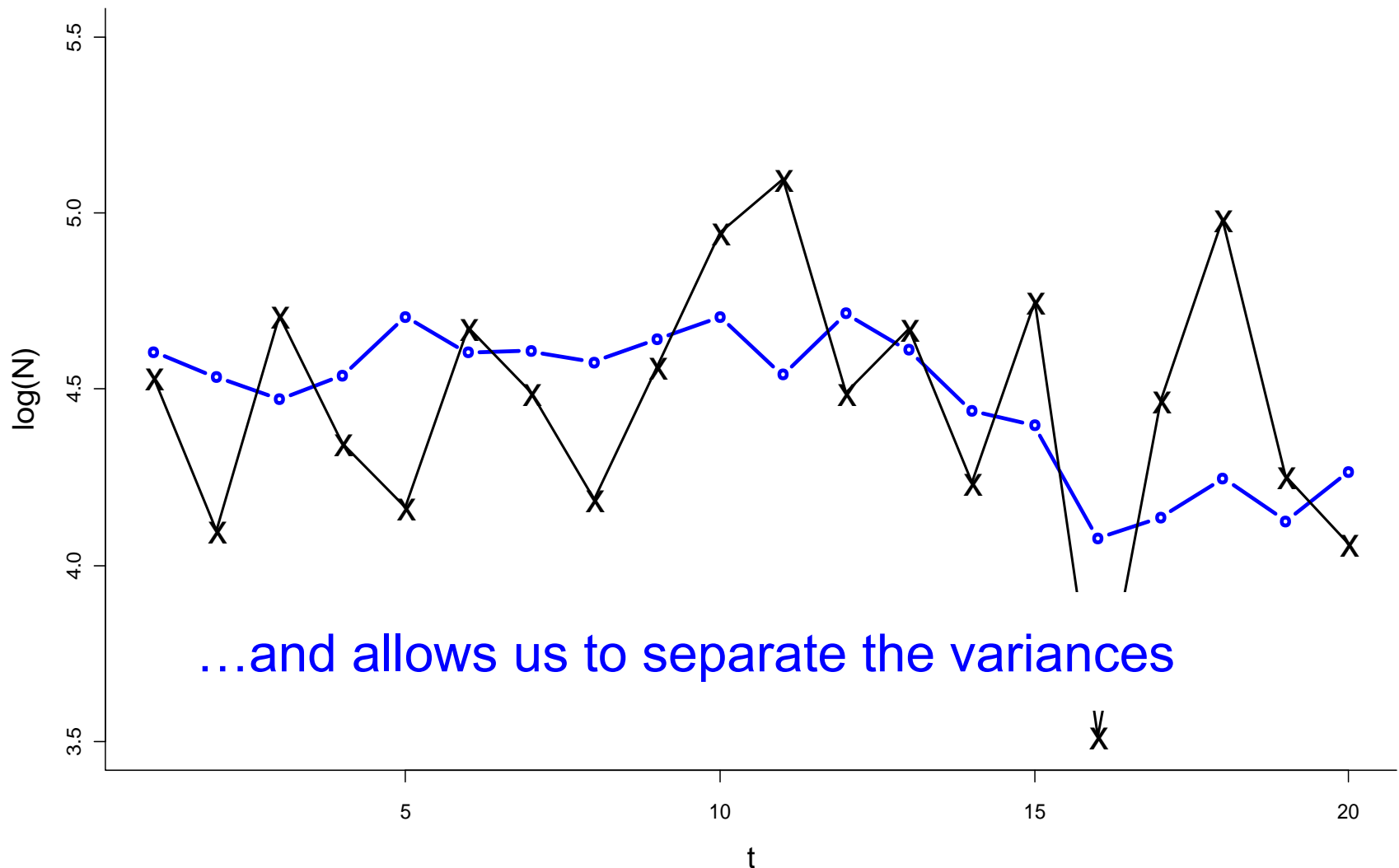
Process error:  $N_t = N_{t-1} \exp(\mu + e_t)$ ,



Observation error:  $N_t = N_{t-1} \exp(\mu)$ ;  $O_t = N_t \exp(\eta_t)$ ;



A state-space model combines a model for the hidden AR-1 process with a model for the observation process



# To fit this model, we have to write it mathematically

Population growth

$$x_t = \log(N_t)$$

$$x_t = x_{t-1} + u + w_t$$

$$w_t \sim \text{Normal}(0, q)$$

$N_t$  is population size

Exponential growth  
model

Normally distributed  
process errors

Observations

$$y_t = \log(O_t)$$

$$y_t = x_t + v_t$$

$$v_t \sim \text{Normal}(0, r)$$

Log-normally  
distributed  
observation errors

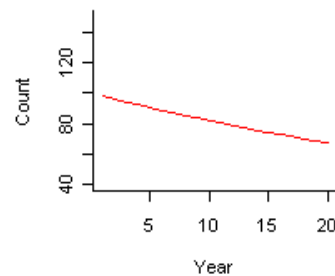
# Let's simulate and try fitting some models

- Open up R and follow after me
- `Lecture_2_univariate_example_1.R`
- `Lecture_2_univariate_example_2.R`
- `Lecture_2_univariate_example_3.R`

# Deterministic, vs. obs. error, vs. proc. error

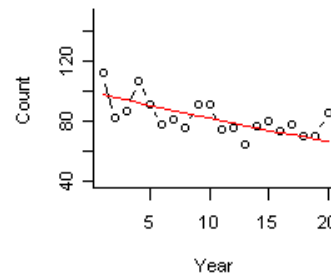
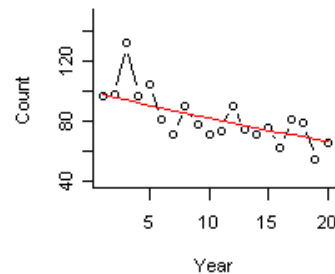
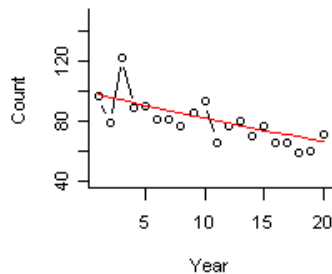
## An example using population decline

a deterministic 2% per year decline



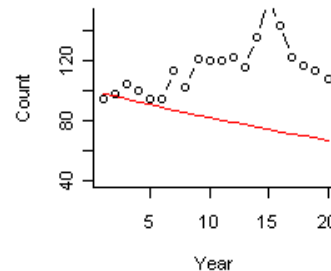
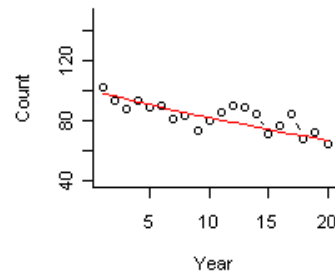
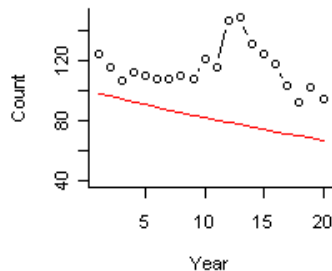
Every year,  
decline 2%

observation error on top of 2% per year decline



Every year,  
decline 2% but  
there is  
observation error

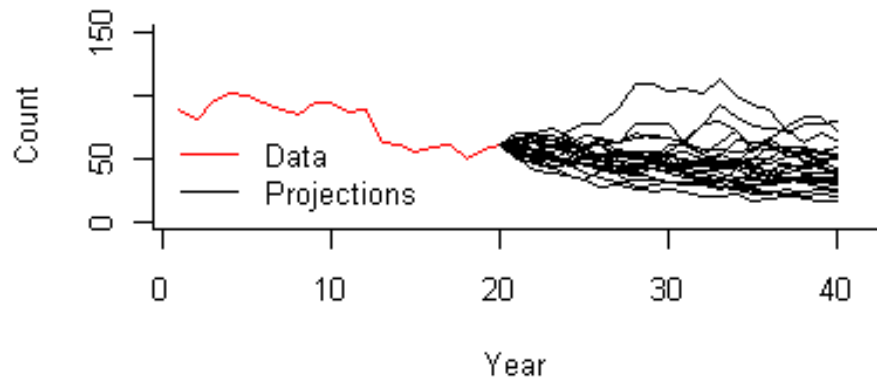
AVERAGE 2% per year decline with year-to-year variation



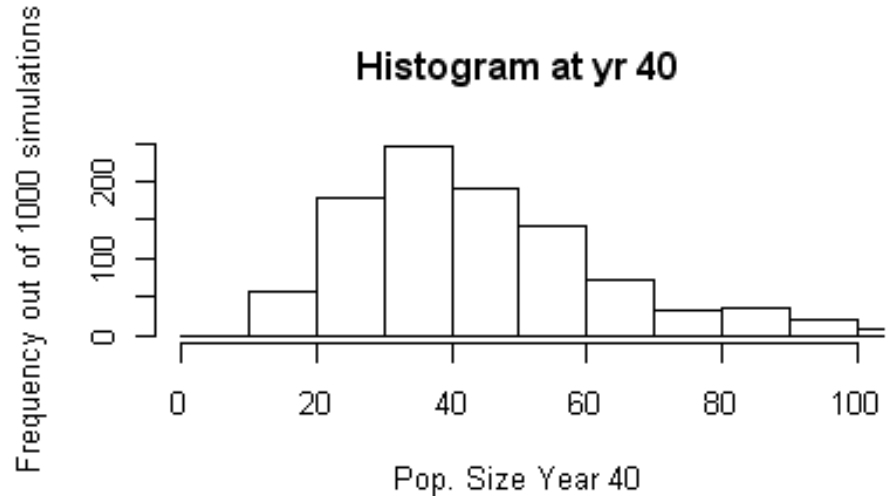
Average yearly  
decline is 2%, but  
actual declines  
vary from year to  
year

# How you model your population data has a large impact on projection of the process

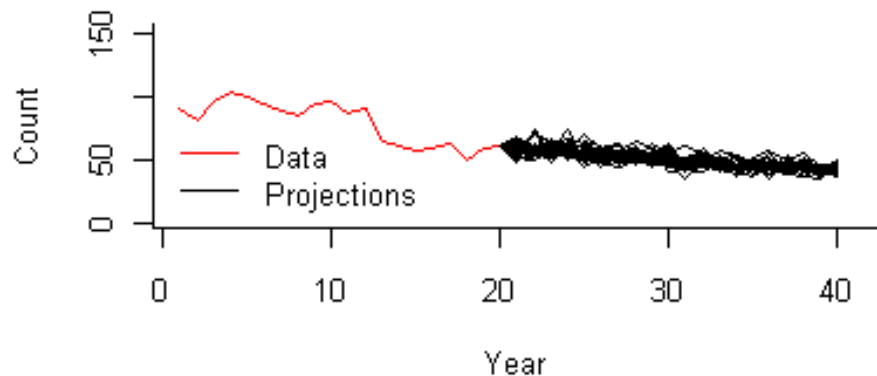
**Process error only**



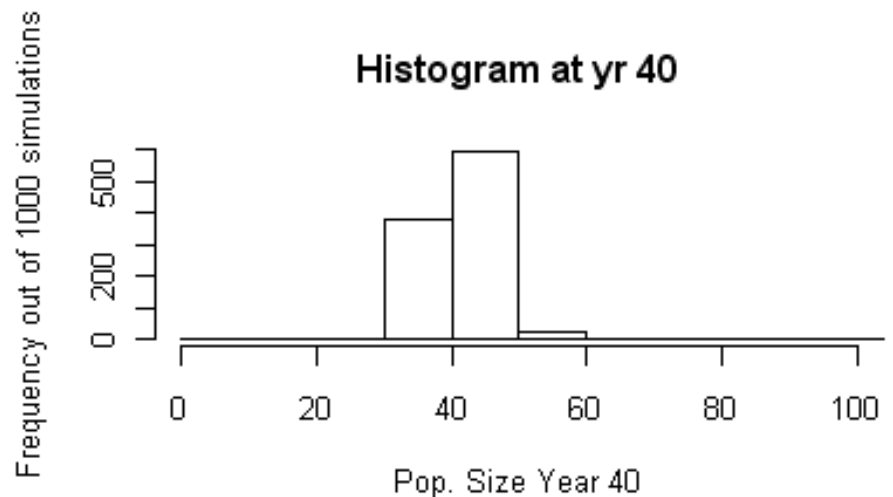
**Histogram at yr 40**



**Observation error only**



**Histogram at yr 40**



# How do you know when to use a process error or observation error model?

- If your time-series data contain both types, use a model with both types unless you know observation error is low.
- To estimate both variances, you need a) 20+ time steps or b) multi-site data.
- If you don't have enough data, you need to use assumptions about one of the variances. Meaning a) fix the value or b) incorporate a prior.
- Diagnostics: Observation error induces autocorrelation in the noise of an autoregressive process. Fit a process-error only model ( $R=0$ ) and check for autocorrelation of residuals (cf. Dennis et al. 1991).

# It's really not “observation error”. It is “non-process” error

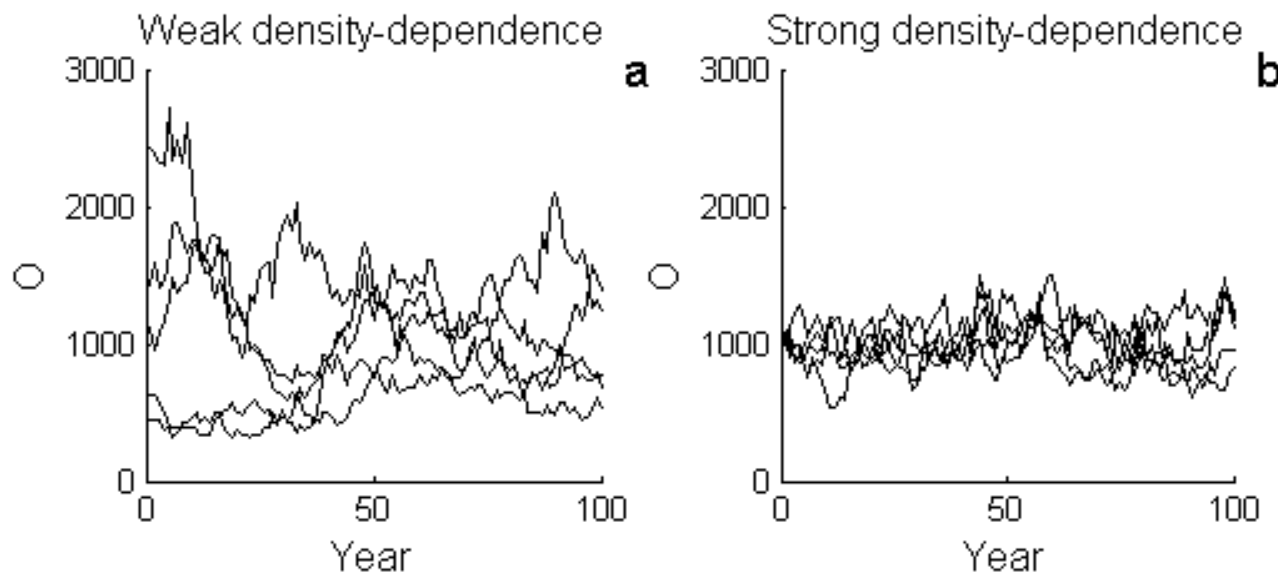
- It's temporally uncorrelated white noise
  - bounces back and forth across some smoother autoregressive trajectory
- Lots of biological processes also create noise that looks like that
  - age-structure cycles
  - density-dependence
  - cyclic variability in fecundity
  - predator-prey interactions
- If your model cannot accommodate that cycling,
  - it tends to get 'soaked' up in the 'non-process' error component
- If your model can accommodate that cycling,
  - estimation of 'observation error' variance can be confounded, unless you have long, long datasets or replicates

# State-space model with density-dependence termed 'mean-reverting'. ---Day 3---

$$N_t = \exp(u + e_t) N_{t-1}^b$$

→  $x_t = b x_{t-1} + u + e_t$  Log-space

$$e_t \sim \text{Normal}(0, q)$$



$b < 1$ : Gompertz density-dependent process

# Computer labs

from the MARSS User Guide

## Chapter 6: Count-based population viability analysis (PVA) using corrupted data

```
library(MARSS)
```

```
RShowDoc("Chapter_PVA.R", package="MARSS")
```

## Chapter 10: Analyzing noisy animal tracking data

```
RShowDoc("Chapter_AnimalTracking.R", package="MARSS")
```